DOCUMENT RESUME

ED 139 628 SE 022 355

AUTHOR Yeany, Russell H., Jr.
TITLE Evaluating Science Teacher Preparation Programs by Assessing Teachers and Pupils.
PUB DATE Mar 77
NOTE 13p.; Paper presented at the annual meeting of the National Association for Research in Science Teaching (50th, Cincinnati, Ohio, March 22-24, 1977); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Evaluation; *Evaluation Methods; Higher Education; *Program Evaluation; *Science Education; Science Teachers; Student Evaluation; *Teacher Behavior; *Teacher Education; Teacher Evaluation
IDENTIFIERS *Learner Outcomes; Research Reports

ABSTRACT

Presented is a program evaluation scheme which provides an opportunity to assess the effectiveness of training program variations which range from pre-training performance of teachers to learning in pupils. The concept of using standardized gain scores is offered as a means of comparing pupil learning when teachers are teaching toward different sets of objectives. (Author/MH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139628

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATING SCIENCE TEACHER PREPARATION PROGRAMS BY ASSESSING TEACHERS AND PUPILS

Russell H. Yeany Jr.
Department of Science Education
University of Georgia

A paper presented at the Annual Meeting of the
National Association for Research
in Science Teaching

Cincinnati, Ohio
March 23, 1977

SE 022 355

EVALUATING SCIENCE TEACHER PREPARATION PROGRAMS
BY ASSESSING TEACHERS AND PUPILS

> The point of preservice teacher education is to produce a social advantage or a savings....
>
> ...pre-service programs should reduce the cost of on-the-job training....
>
> ...should lead to greater teacher productivity of desirable pupil outcomes than the alternative. (Turner, 1971, p. 10)

The instructional techniques and materials employed in science teacher preparation activities usually have the expressed intent of modifying the teachers' behavior. The inference then is that when these behaviors are used in the science classroom, they will improve pupil performance.

An early stage in the teacher training process can be identified as the competency acquisition phase. In this stage the teacher is able to recognize a competency and begins to internalize it into his own cognitive structure which later will influence his teaching behavior.

A later stage is the skill application phase where a teacher exhibits an overt behavior which is largely influenced by the level of competency acquisition. For example, teachers can study and apply the various levels of the Teaching Strategies Observation Differential (a classroom observation system) until they can recognize and categorize different teaching styles and communicate within the language of the system. Later, they can manipulate their own teaching strategies in relation to the teaching/learning environment and through skill application exhibit a higher level of competency acquisition.

Historically, teacher training programs have been judged on the basis of the acquisition of knowledge with limited demonstration of skill exhibition. Only recently has the use of pupil outcomes been promoted as measures of

training program effectiveness (e.g., Dunkin and Biddle, 1974; Popham, 1971). McNeil and Popham (1973) advocated that the ultimate criterion of teachers' competence be their impact upon learners. This is also the position taken by Okey (1977) in a companion paper to this report.

Science teacher program personnel need to concern themselves at all levels of skill acquisition and application and with resultant pupil changes. The purpose of this paper is to promote this awareness and offer a mechanism for carrying out program research and evaluation within the above framework.

The Assessment of Skill Acquisition

Research and evaluation methods related to skill acquisition revolve around three generic stages of activities: 1) the pre-treatment or pre-instruction assessment of the skill and related variables, 2) involvement in instructional activities designed to enhance the competency, and 3) post-assessment of skill acquisition or improvement and related variables (see Figure 1).

Collecting pre-instruction data can serve different purposes. First, the information may be used, in the absence of a control or comparison group, as base line data to determine the effects of training. Second, some pre-data may be useful as predictors of success in relation to such variables as personality type or readiness level. Such information can be applied in a prescriptive format where instructional resources are matched to the situation of highest probable impact.

During research and evaluation efforts on skill acquisition, the instructional format may be varied more than normal in order to provide an opportunity to compare training modes. For example, the comparison might be between a self-paced modular format and an instructor-centered lecture/ discussion format related to the same skill, or among self study, peer

Figure 1. Assessing the acquisition of teaching skills.

Figure 2. Assessing the application of teaching skills.

discussion format related to the same skill, or among self study, peer participation or field experience with pupils as a means of acquiring a teaching skill.

Post-assessment of skill acquisition may encompass three types of teacher measures: 1) cognitive, 2) affective, and 3) skill exhibition. Cognitive performance and attitude data can normally be collected through conventional pencil and paper measures. But, in many instances skill exhibition data must be collected through visual or auditory observation and coded with the aid of an analysis scheme to facilitate communications and comparisons (Yeany, 1977 and Yeany and Capie, 1977).

Skill Application

Because the modification of pupil behavior is a terminal goal of teaching, one measure of teaching effectiveness is made through an assessment of the change brought about in pupils during the skill application phase. Training program evaluation should not end with an assessment and judgment related to the level of skill acquisition without determining the degree to which teachers are able to apply a competency and modify pupil behavior.

The Skill Application phase of training research and evaluation involves three related procedures: 1) the collection of pre-application pupil data, 2) the application of the skill in an instruction setting and, 3) post-application data collection to assess cognitive gains and the attitudes of pupils toward the content and methods of instruction (see Figure 2). The acid-test of skill application is whether a desired change in pupil behavior can be brought about when the conditions are right and instruction is designed and applied to do so. In other words, is the teaching effective?

To make decisions about training program effectiveness based on pupil performance is to base decisions on data twice removed from the source of influence (i.e., the level of skill acquisition and application both act as filters). But, the ultimate influence of training programs is meant to be exerted on pupil performance through the trainees.

There are at least two ways that skill application evaluation can be carried out. A fairly recent model has been developed by Popham, Baker, Millman and McNeil (1972). These authors describe the teaching performance test (minilesson) as a means of analyzing and assessing teaching effectiveness. In this procedure, the teacher is given an explicit, measurable objective suitable for a short lesson. A sample test item is provided to clarify the objective, and the teacher prepares and teaches a lesson designed to accomplish the objective. At the end of the lesson a cognitive and affective post-test is administered to the pupils. On the basis of these, a judgment of teaching effectiveness can be made. And, on the basis of such data from all program trainees, a judgment can be made about program effectiveness. If random sets of trainees have experienced different treatments, the pupil performance data can be statistically analyzed to measure relative efficacy of training methods designed to bring about teacher behaviors that influence pupil performance.

The minilesson approach as suggested by Popham, et al. (1972) is an appealing and viable means of evaluating teacher preparation programs and has been used for such purposes (e.g., Rezba, Lahnston and Lapp, 1975). But, the skill application phase of program evaluation should not be limited to this procedure. An attempt should be made to assess effects of training on the acquisition of a teaching skill in a less contrived and restricted context. Also, the selection of the objectives should be the perogative of the

teacher and should be permitted to vary with the immediate instructional content in order to increase the generalizability across a sizable set of objectives. On the surface, this may appear to be a bold suggestion. But, when the dependent variable is pupil performance as influenced by teaching toward one externally imposed objective, the judgment about training program effectiveness must be very tenative, to say the least. The only criterion which should be imposed on the selection of an objective is whether teaching toward it will provide an opportunity to apply the teaching competency which is being assessed.

If absolute judgments about the teaching effectiveness of each individual in the program are to be made, the above suggestion presents few problems. Simply, decide what kinds of post-test scores are acceptable and base the decision on that point. On the other hand, if relative comparisons are to be made (e.g., comparing the training modes of self study, peer practice and field based practice), the dependent variable of pupil performance needs to be redefined.

The concept of using standardized gain scores is offered here as a means of comparing pupil outcomes when teachers are teaching toward different sets of objectives. This procedure necessitates the administration of identical or parallel pre-tests and post-tests for each set of objectives. The objectives are used to provide a focus for the development of the tests.

To carry out the analyses, the dependent variable is redefined as the amount of change in average pupil performance from pre to post-test in coded standard deviation units. Each pre-test is used as normative data for the post-test. The raw pre-test scores are standardized to T-scores with a mean of 50 and a standard deviation of 10 where $T = 10z + 50$ and $z = \frac{x-\bar{x}}{s}$ (Glass and Stanley 1970). The raw class mean score on the post-test is then converted

to a standard score by finding the difference between it and the raw pre-test mean and dividing that value by the standard deviation of the pre-test and then converting to a T-score as follows:

$$Z = \frac{X_{pre} - X_{post}}{S_{post}} \qquad \text{and} \qquad T_{post} = 10z + 50.$$

It should be evident that the generation of the T pre-score is not computationally necessary because the mean of these scores will always equal 50 with s = 10. But, it may be a necessary conceptual step to realize that the difference between 50 and the T_{post} - score represents the gain in pupil achievement in one-tenth standard deviation units (i.e., an average T_{post} -score of 60 represents one full standard deviation gain in class achievement. It should also be evident that the z-score could be used but negative gain scores would be a problem.

The average class T_{post} -scores are then treated as the dependent variables with conventional statistical analysis procedures to determine significant differences in pupil outcomes which can be attributed to variation in training program format. If the research/evaluation design being employed makes use of randomization across comparison groups, the threat to internal validity inherent in the use of a pre-test will not be present. If the threat is present, it is suggested that the pre-test be administered to a random one-half of the pupils for the normative purposes and the remaining one-half of the post-tests be used to assess achievement gain.

A comprehensive evaluation of training materials and activities should include assessment in both competency acquisition and application (Figure 3). A failure to do both (for some skills) demonstrates an incongruency between terminal goals (i.e., the modification of pupil behavior) and the evaluation

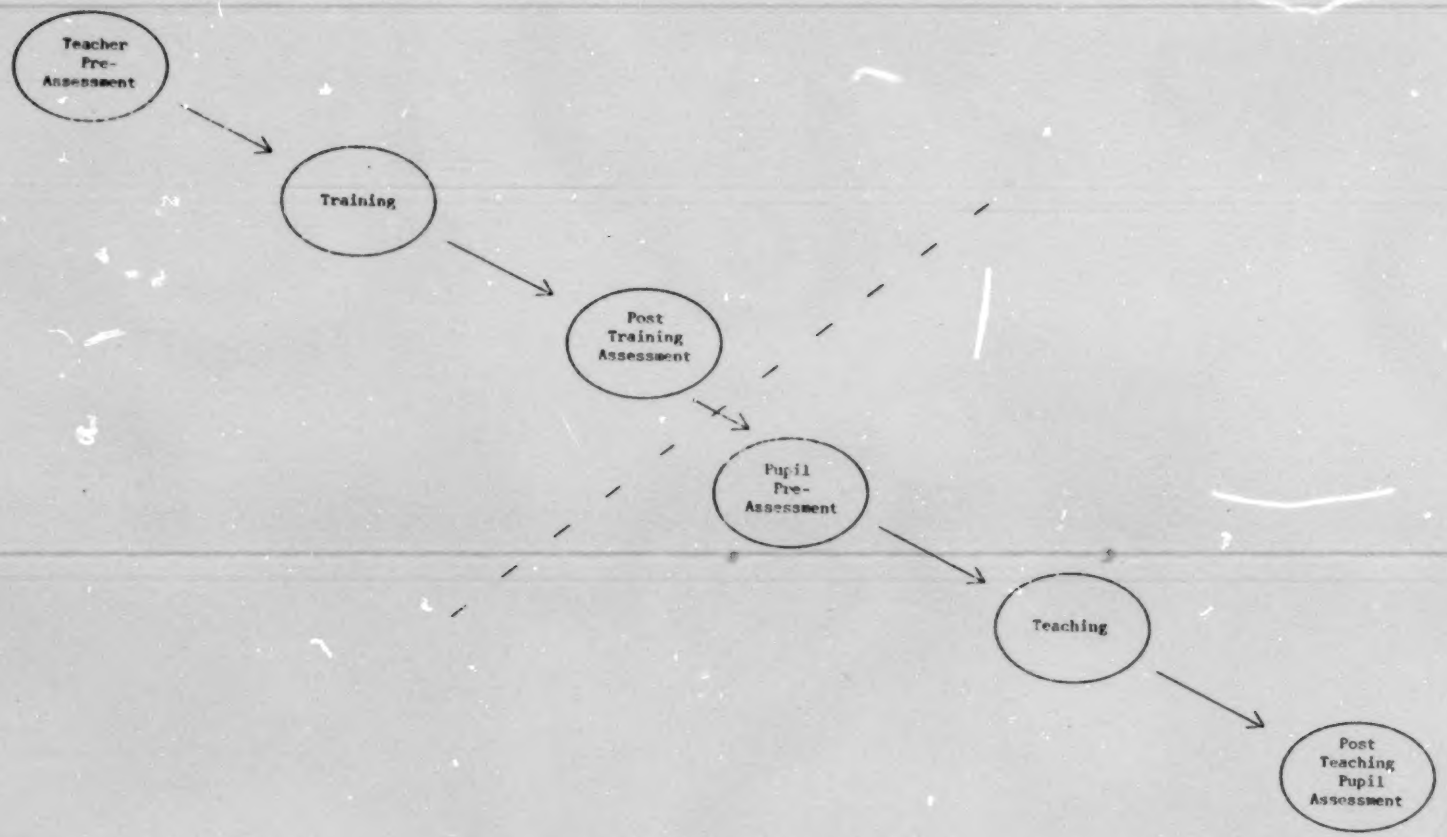

Figure 3. Assessing Teaching Skill Acquisition and Application.

scheme; or, it denies that the real goal in teacher training programs is to influence teaching and learning in the science classroom. Also, the setting and selection of content and resultant outcomes should be free to vary in order to allow greater generalizability of the findings. To do so will reduce the tentative nature of our training program decisions, and perhaps, produce a social advantage and reduce the "cost" of on-the-job training.

REFERENCES

Dunkin, M. J. and Biddle, B. J. The Study of Teaching. New York: Holt, Rinehart and Winston, Inc., 1974.

Glass, G. V, and Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

McNeil, J. D. and Popham, W. J. The assessment of teacher competence. In R. M. Travers (Ed.), Second Handbook of Research in Teaching. Chicago: Rand McNally, 1973.

Okey, J. R., Using learner performance to measure science teacher effectiveness, a paper presented at the meeting of the National Association for Research in Science Teaching, Cincinnati, March 1977.

Popham, W. J. Performance tests of teaching proficiency: Rationale, development and validation. American Educational Research Journal, 8(1), January, 1971.

Popham, W. J., Baker, E., Millman, J., and McNeil, J. D. Teaching Improvement Kit. Los Angeles: Instructional Appraisal Services, 1972.

Rezba, R., Lahnston, A., and Lapp, D. Evaluating teacher and program effectiveness through pupil performance measures, a paper presented at the meeting of the National Association for Research in Science Teaching, San Francisco, April 1976.

Turner, R. L. Conceptual foundations of research in teacher education. In Smith, O. B. (Ed.) Research in Teacher Education: A Symposium. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971.

Yeany, R. H. The Effects of Model Viewing with Systematic Strategy Analysis on the Science Teaching Styles of Pre-Service Teachers, Journal of Research in Science Teaching, (in press).

Yeany, R. H. and Capie, W. R. An analysis system for describing strategies of teaching data manipulation, a paper presented at the meeting of the Southeastern Association for the Education of Teachers in Science, Chattanooga, 1977.